



TITLE:

Dynamics and mechanisms of clonal expansion of HIV-1-infected cells in a humanized mouse model

AUTHOR(S):

Satou, Yorifumi; Katsuya, Hiroo; Fukuda, Asami; Misawa, Naoko; Ito, Jumpei; Uchiyama, Yoshikazu; Miyazato, Paola; ... Bangham, Charles R. M.; Koyanagi, Yoshio; Sato, Kei

CITATION:

Satou, Yorifumi ...[et al]. Dynamics and mechanisms of clonal expansion of HIV-1-infected cells in a humanized mouse model. Scientific Reports 2017, 7: 6913.

ISSUE DATE:

2017-07-31

URL:

<http://hdl.handle.net/2433/227890>

RIGHT:

© The Author(s) 2017.; This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

SCIENTIFIC REPORTS

OPEN

Dynamics and mechanisms of clonal expansion of HIV-1-infected cells in a humanized mouse model

Yorifumi Satou^{1,2,3}, Hiroo Katsuya^{1,2}, Asami Fukuda^{1,3}, Naoko Misawa⁴, Jumpei Ito⁵, Yoshikazu Uchiyama⁶, Paola Miyazato^{1,2}, Saiful Islam^{1,2}, Ariberto Fassati⁷, Anat Melamed⁸, Charles R. M. Bangham⁸, Yoshio Koyanagi⁴ & Kei Sato^{4,9}

Combination anti-retroviral therapy (cART) has drastically improved the clinical outcome of HIV-1 infection. Nonetheless, despite effective cART, HIV-1 persists indefinitely in infected individuals. Clonal expansion of HIV-1-infected cells in peripheral blood has been reported recently. cART is effective in stopping the retroviral replication cycle, but not in inhibiting clonal expansion of the infected host cells. Thus, the proliferation of HIV-1-infected cells may play a role in viral persistence, but little is known about the kinetics of the generation, the tissue distribution or the underlying mechanism of clonal expansion *in vivo*. Here we analyzed the clonality of HIV-1-infected cells using high-throughput integration site analysis in a hematopoietic stem cell-transplanted humanized mouse model. Clonally expanded, HIV-1-infected cells were detectable at two weeks post infection, their abundance increased with time, and certain clones were present in multiple organs. Expansion of HIV-1-infected clones was significantly more frequent when the provirus was integrated near host genes in specific gene ontological classes, including cell activation and chromatin regulation. These results identify potential drivers of clonal expansion of HIV-1-infected cells *in vivo*.

Human immunodeficiency virus type 1 (HIV-1) is an exogenous retrovirus with worldwide distribution. HIV-1 infects CD4⁺ cells such as macrophages and helper T cells and causes acquired immunodeficiency syndrome (AIDS). Combination anti-retroviral therapy (cART) potently controls HIV-1 replication and prevents the onset of AIDS in most treated individuals¹. However, cART does not eradicate the virus: a reservoir of HIV-1-infected cells persists despite prolonged therapy², and infected individuals cannot interrupt treatment³. Furthermore, persistent HIV-1 infection causes various complications in infected individuals, such as HIV-associated neurocognitive disorders and a high incidence of certain cancers^{4,5}.

Since HIV-1 is a retrovirus, the viral RNA genome is reverse transcribed into a double stranded DNA. The viral DNA is integrated into the host genomic DNA to form a provirus, which is subsequently transcribed to generate new viral particles. In the absence of cART, there is active viral replication, resulting in a high viral load in the plasma. When infected individuals are treated with cART, the plasma viral RNA typically becomes undetectable by standard assays. In this situation, HIV-1 proviral DNA contributes to the maintenance of the viral reservoir, and hence much effort has been recently dedicated to study the distribution of HIV-1 proviral integration sites and the mechanisms of proviral reactivation^{6–8}.

Previous studies have shown that HIV-1 is preferentially integrated into gene bodies with active transcription in *in vitro* infection^{9,10}. Recent studies of HIV-1 provirus from *ex vivo* samples of infected individuals have

¹International Research Center for Medical Sciences (IRCMS), Kumamoto University, Kumamoto, 860-0811, Japan. ²Center for AIDS Research, Kumamoto University, Kumamoto, 860-0811, Japan. ³Priority Organization for Innovation and Excellence, Kumamoto University, Kumamoto, 860-0811, Japan. ⁴Laboratory of Systems Virology, Institute for Frontier Life and Medical Sciences, Kyoto University, Kyoto, 606-8507, Japan. ⁵Division of Human Genetics, Department of Integrated Genetics, National Institute of Genetics, Shizuoka, 411-8540, Japan. ⁶Department of Medical Physics, Faculty of Life Sciences, Kumamoto University, Kumamoto, 862-0976, Japan. ⁷Division of Infection and Immunity, University College London, London, WC1E 6BT, United Kingdom. ⁸Department of Immunology, Division of Infectious Diseases, Imperial College London, London, W2 1PG, United Kingdom. ⁹CREST, Japan Science and Technology Agency, Saitama, 322-0012, Japan. Yorifumi Satou and Hiroo Katsuya contributed equally to this work. Correspondence and requests for materials should be addressed to Y.S. (email: y-satou@kumamoto-u.ac.jp)

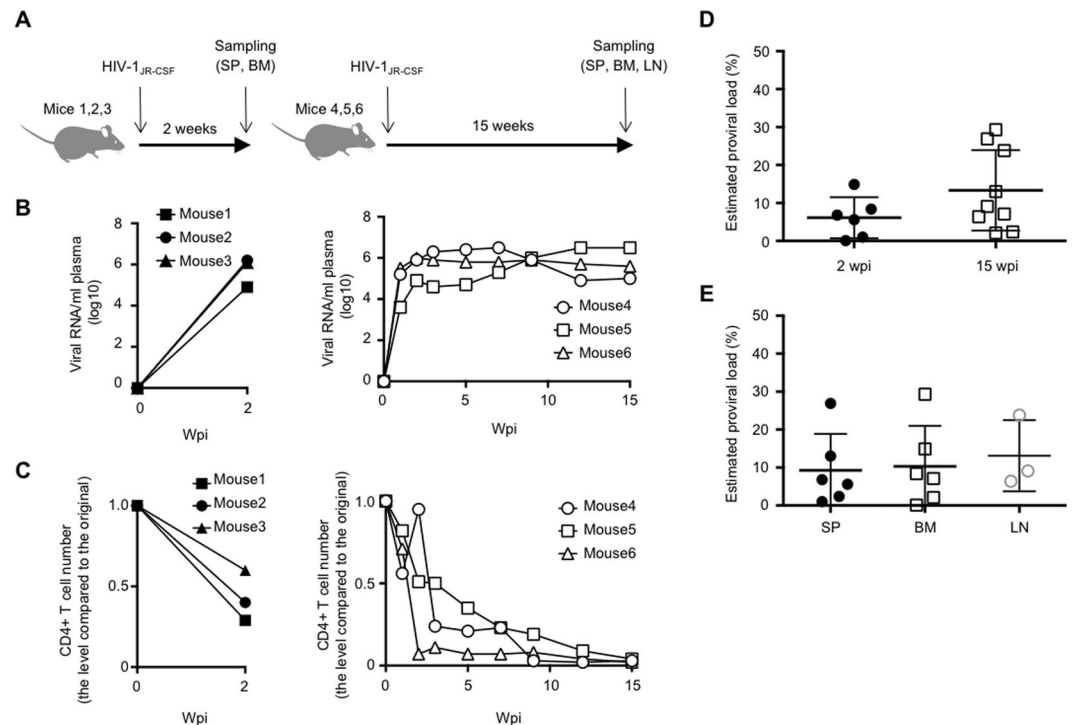


Figure 1. HIV-1 infection in the humanized mice. **(A)** Diagram showing the experimental design. Bone marrow (BM) and spleen (SP) were collected from 2 wpi mice and BM, SP, and lymph node (LN) were collected from 15 wpi mice for clonality analysis. Longitudinal analysis of plasma viral load **(B)** and CD4⁺ T-cell number **(C)** in 2 wpi and 15 wpi humanized mice. **(D)** Comparison of proviral load between 2 wpi and 15 wpi mice. **(E)** Comparison of proviral load among different tissues, SP, BM, and LN.

revealed several additional key observations. First, the strong selection pressure exerted on HIV-1 by escape from anti-retroviral drugs or anti-viral immunity, resulting in a high frequency of defective proviruses *in vivo*^{11–13}. Second, it has been reported recently that HIV-1-infected cells can undergo clonal expansion, as observed in infection with the related retrovirus human T-cell leukemia virus type 1 (HTLV-1)^{12, 14, 15}. The clonal abundance of infected cells can be quantified by using integration site analysis^{16, 17}. HIV-1 integration was frequently detected near cancer-related genes in the expanded clones of infected individuals^{14, 15}. Clonal expansion may contribute to the maintenance of the HIV-1 reservoir^{13, 18}, but the precise kinetics of generation of HIV-1-infected clones and the underlying mechanism of clonal expansion *in vivo* are still poorly understood.

In vitro studies have afforded much progress in HIV-1 research, including the development of anti-retroviral drugs. However persistent HIV-1 infection *in vivo* is associated with active host cell dynamics, i.e. differentiation, activation, quiescence, and homeostasis of CD4⁺ T cells, limiting the usefulness of *in vitro* cell culture systems¹⁹. Direct analysis of clinical samples is complicated by the wide variation between individuals in the time since initial infection, the genotype of the founder virus, and the immunological status of the host. To understand HIV-1 persistence in the host, *in vivo* models of infection are therefore required. Humanized mice transplanted with human CD34⁺ hematopoietic stem cells have been shown to be useful to study HIV-1 infection *in vivo*^{20–22}. One can study CD4⁺ T-cell dynamics in the humanized mouse model, including generation, differentiation, activation, and homeostasis²³.

Here, we investigated whether the humanized mouse model recapitulates clonal expansion of HIV-1 infected cells as detected in patients, and whether it can provide mechanistic insight into the process of clonal selection of virus-infected cells. We found rapid generation of a large number of clones of HIV-1-infected cells and proliferation of individual clones within 2 weeks of infection. This period corresponds to the early stage of infection in human, before the emergence of the host immune response. Furthermore, in the expanded clones we found significant enrichment of HIV-1 proviruses integrated near genes involved in lymphocyte activation and chromatin regulation, providing the evidence for underlying mechanism of clonal expansion of HIV-1-infected cells.

Results

HIV-1 infection in the humanized mice: plasma viral load, CD4⁺ T-cell count and proviral load. We transplanted human CD34⁺ hematopoietic stem cells into the liver of new-born recipient NOD/SCID/*Il2r^{null}* mice (NOG mice). After 10–12 weeks post-transplantation, the humanized mice (n = 6) were intraperitoneally infected with the CCR5-tropic HIV-1_{JR-CSF}. We divided the six mice into two groups; respectively 2 weeks post infection (wpi) and 15 wpi (Fig. 1A and Supplementary Table S1). The plasma viral RNA was around 10⁶ copies/mL after 1–2 wpi and this high viremia was still present at 15 wpi (Fig. 1B). The CD4⁺ T-cell count fell to half the initial level by 2 wpi and to less than 5 percent of the initial level by 15 wpi (Fig. 1C), consistent with our

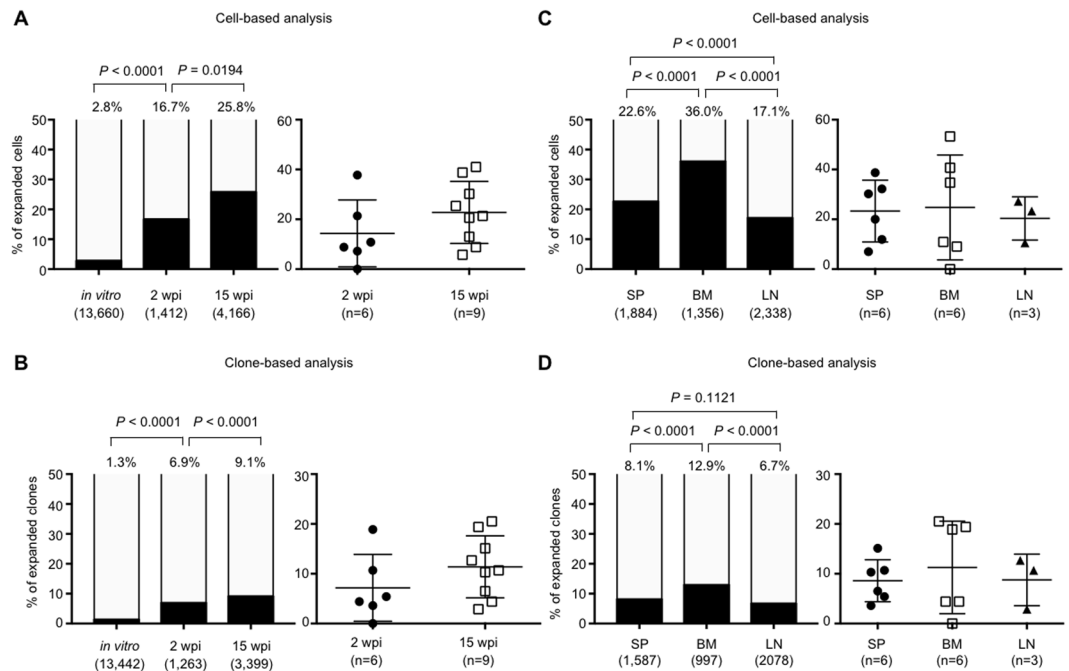


Figure 2. Cell-based and clone-based integration site analysis. (A) The percentages of expanded cells within the total number of infected cells were analyzed using genomic DNAs from tissues of 2 wpi or 15 wpi mice. DNA extracted from Jurkat cells infected with HIV-1 *in vitro* was analysed as a control. Cumulative data are shown in the left panel. The value for each individual mouse is shown in the right panel. (B) Clone-based analysis of 2 wpi or 15 wpi mice. (C) Comparison of the degree of clonal expansion at different tissues. Cumulative results of cell-based analysis are shown in the left panel, and the value of each individual mouse is shown in the right panel. (D) Clone-based analysis is also shown. The number of infected cells or clones in each sample is shown under the graph of the left panel. The number of mice or tissues analysed is shown under the graph of the right panel.

previous report²⁴. To analyze HIV-1 proviruses, we extracted genomic DNA from spleen (SP) and bone marrow (BM) of 2 wpi mice. We also extracted DNA from lymph node (LN) of 15 wpi mice as well as SP and BM. The estimated proviral load (PVL) was much higher than that typically observed in peripheral blood mononuclear cells (PBMCs) of infected individuals (Fig. 1D)²⁵, and there was no significant difference in PVL between different tissues (Fig. 1E). We also observed wide variation in estimated PVL levels among individual mice and different tissues (Fig. 1D and E).

Clonality of HIV-1 infected cells in infected humanized mice. To investigate if HIV-1 infection resulted in clonal expansion of HIV-1-infected cells in the humanized mice, we mapped and quantified HIV-1 integration sites by using linker-mediated PCR as previously described¹⁶, with minor modifications (Supplementary Fig. S1). We also calculated clonal abundance using information on the DNA shear site¹⁷. In total, we identified 4,662 and 13,442 unique HIV-1 integration sites in humanized mice and in *in vitro* infection, respectively. Here, we defined an infected clone that was detected more than once in the same sample as an “expanded clone”, whereas an infected clone detected only once was called a “singleton clone”. We found that the extent of clonal expansion of the infected cells in 2 wpi mice was much greater than that of cells infected *in vitro* ($P < 0.0001$, Fig. 2A). In 2 wpi mice, 87 clones (236 cells) were expanded, whereas 1,176 clones (1,176 cells) were identified as singleton clones. In 15 wpi mice, we identified 309 expanded clones (1,076 expanded cells) and 3,090 singleton clones (3,090 single cells) (Fig. 2A and B). The proportion of expanded clones increased significantly from 2 wpi to 15 wpi, as judged by both cell-based and clone-based analysis ($P = 0.0194$ and $P < 0.0001$, respectively, Fig. 2A and B). Clonal expansion was most evident in BM, although there was wide variation between individual mice (Fig. 2C and D). The relative clonal abundance of each infected clone is shown in Fig. 3A and B. The expanded clones represented between 7.6% and 34.7% of total infected cells in 2 wpi mice (except for one BM tissue with very low PVL, 0.1%) and between 10.6% and 53.3% in 15 wpi mice. These results indicate that HIV-1 infected cells underwent clonal expansion within 2 weeks in the humanized mouse model.

HIV-1 preferentially integrated into the gene body of highly-expressed host genes in the humanized mice. It has been reported that HIV-1 preferentially integrates into actively transcribed genes in human cells^{9,10}. We next analyzed the distribution of the HIV-1 integration sites (IS) in the humanized mice. Integration sites in cells infected *in vitro* with HIV-1 or human T-cell leukemia virus type 1 (HTLV-1) were analyzed as controls. Consistent with previous reports^{9,26}, HIV-1 preferentially integrated within host genes. The global distribution of HIV-1 integrations *in vitro* was similar to the distribution *in vivo*, whereas HTLV-1 integration *in vitro* did not show such a tendency (Supplementary Fig. S2). Closer analysis showed that HIV-1

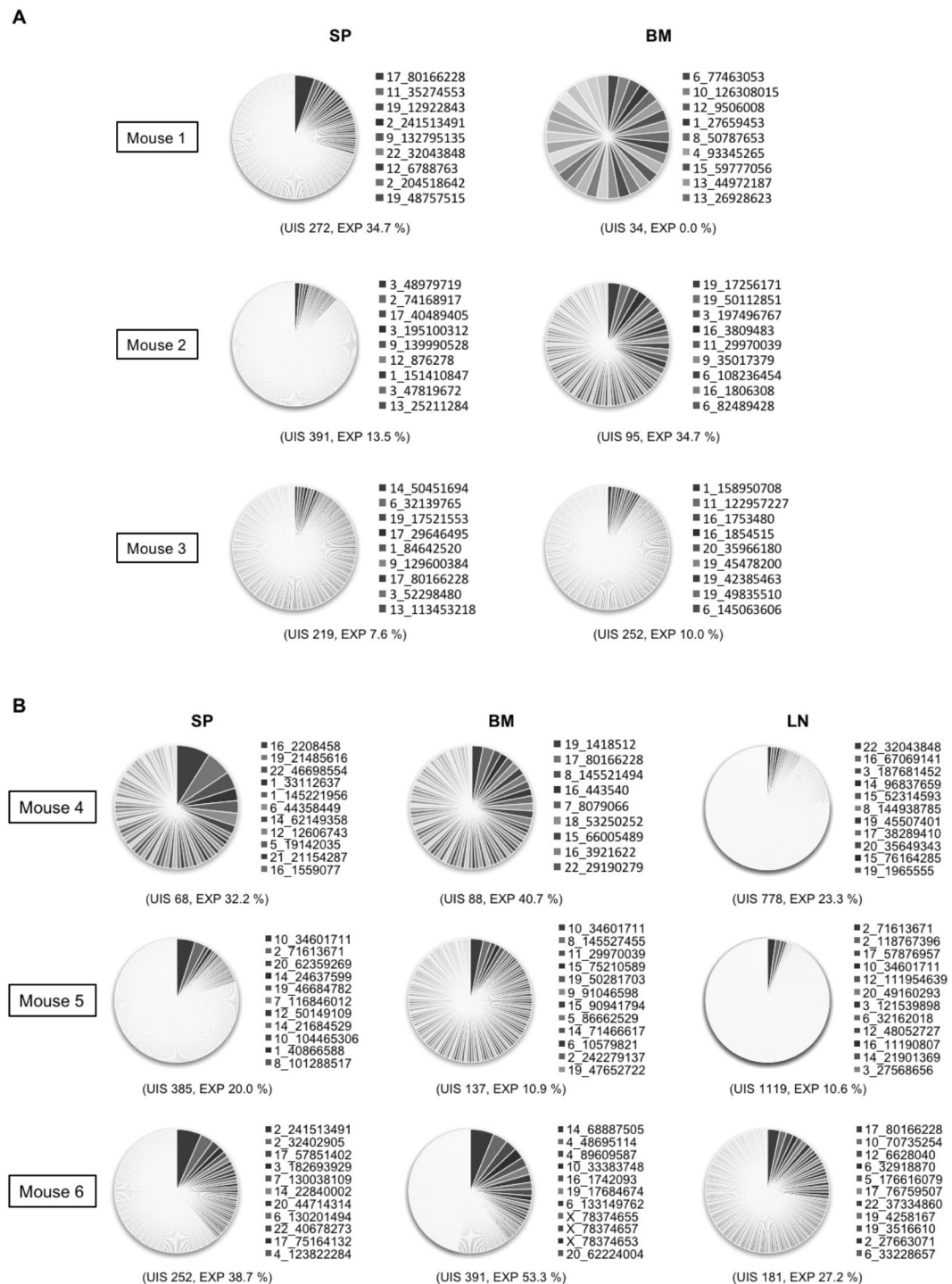


Figure 3. Degree of clonal expansion of HIV-1-infected cells in individual mice. **(A,B)** Each pie chart graph shows the relative abundance of each individual clone in 2 wpi **(A)** and 15 wpi **(B)** mice. The area represents the relative degree of expansion of each individual clone in the tissue sample. DNA samples were analysed according to the experimental procedure shown in Supplementary Figure S1 and described in Methods. UIS: Number of unique integration site, EXP: % of expanded clone. A list of integration sites of the highly-expanded clones is shown on the right side of the pie chart as chromosome, positions.

IS in the humanized mice were frequently located inside the gene body (Fig. 4A). The frequency of integrations within genes was lower than that seen in *in vitro* HIV-1 infection (84.7%), but much higher than that in HTLV-1 infection *in vitro* (51.5%) (Fig. 4A). The frequency of integration within genes was higher in 15 wpi mice than in 2 wpi mice (79.1% and 75.1%, respectively), which reached statistical significance ($P = 0.0033$). There was no significant difference in the frequency of HIV-1 integration within genes between expanded and singleton clones, either in 2 wpi or 15 wpi mice (Fig. 4B). We also analyzed HIV-1 integration into exonic or intronic regions of

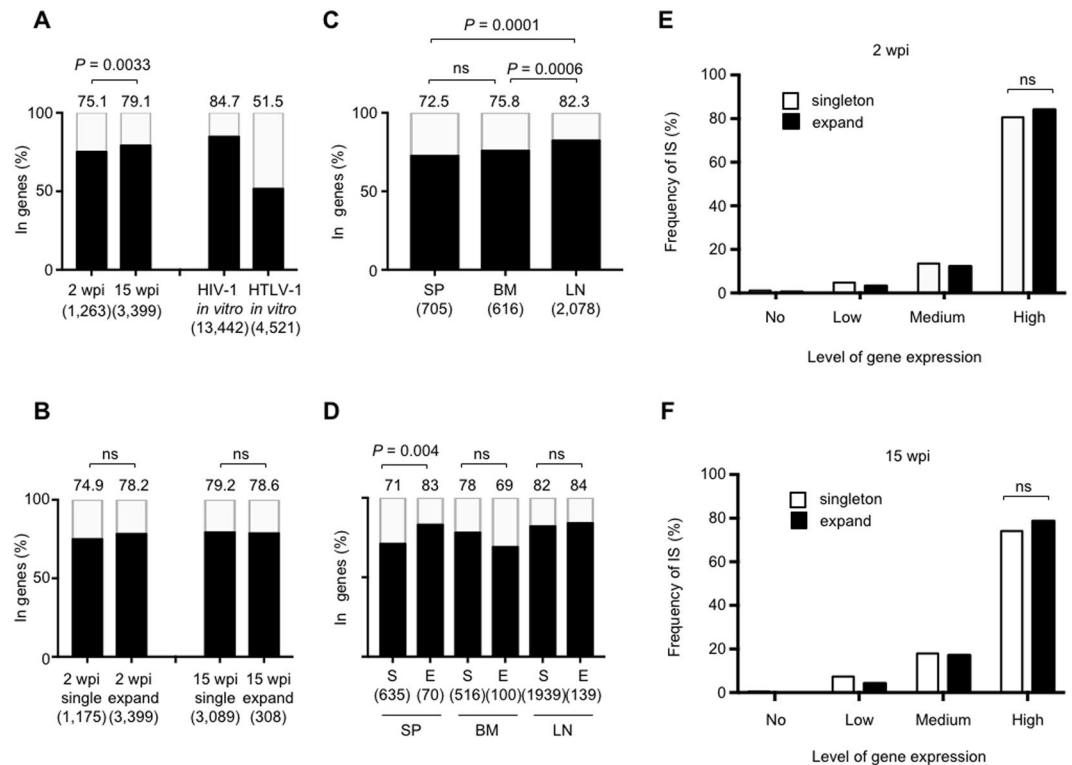


Figure 4. Relationship between HIV-1 integration sites and the host genes. (A) Frequencies of the integration sites located in genes in 2 wpi or 15 wpi mice are shown in the bar graph. Jurkat cells infected with HIV-1 or HTLV-1 *in vitro* were used as controls. (B) Comparison of the frequency of expanded and singleton clone in 2wpi and 15 wpi mice. (C) Frequencies of integration sites within genes are shown for different tissues. (D) Comparison of the frequency between expanded and singleton clones in different tissues. S: singleton clones, E: expanded clones. Numbers above the graph indicate the percentage. The number of infected cells or clones in each sample is shown under the graph. (E,F) Relationship between integration sites and level of expression of the corresponding host genes in 2 and 15 wpi mice. The cumulative results from different tissues of 2 wpi mice (E) and 15 wpi mice (F) are shown in the bar graphs. ns: no significant difference.

host genes, and found a high frequency of HIV-1 integration into intronic regions, as reported previously^{10,12} (Supplementary Fig. S3). There was a significantly increased frequency of HIV-1 integration into gene bodies in LN (82.3%) compared with SP (72.5%) or BM (75.8%) (Fig. 4C). In SP the frequency of integration into gene bodies was significantly higher in expanded clones than in singleton clones ($P = 0.004$, Fig. 4D), but not in BM and LN. There was no significant association between the transcriptional orientation of the HIV-1 provirus and that of the host gene, either in the whole proviral population (Supplementary Fig. S4A and B) or in any tissue examined (Supplementary Fig. S4C and D).

Previous reports have shown that HIV-1 is frequently integrated into host genes that are highly expressed^{9,12}. To investigate whether this is also the case in our model, we first performed RNA-sequencing (RNA-seq) analysis using CD4⁺ T cells from SP of two humanized mice without HIV-1_{JR-CSF} infection to generate average transcription values of each gene, and then compared the distribution of HIV-1 integration sites and the intensity of host gene expression. The results showed frequent integration of HIV-1 into highly expressed genes in both 2 wpi and 15 wpi mice (Fig. 4E and F), consistent with previous studies on PBMCs from HIV-1-infected individuals¹². However, the frequency of integration into highly expressed genes did not differ significantly between singleton and expanded clones, either in 2 wpi or 15 wpi mice (Fig. 4E and F).

Enrichment of HIV-1 integration sites in expanded clones near host genes associated with lymphocyte activation and chromatin regulation. In patients, HIV-1 IS in expanded clones were frequently detected near host genes associated with cell growth^{12,13}, but it is unclear whether this is a general mechanism. To address this issue, we performed gene ontology (GO) enrichment analysis on IS identified in this study. The analysis revealed that, while IS in singletons were enriched in genes associated with host-pathogen interactions (Fig. 5A and Supplementary Fig. S5), in expanded clones, IS were enriched into genes related to lymphocyte activation and chromatin regulation (Fig. 5B, Supplementary Fig. S6, and Supplementary Tables S2 and S3). The data from cells infected with HIV-1 *in vitro* showed a tendency similar to that observed in singleton clone (Fig. 5C). Comparisons between each enriched GO terms are shown in Fig. 5D. There were overlaps between singleton clones and *in vitro* infection; and between singleton and expanded clones, but not between *in vitro* infection and expanded clones when we analyzed the top 14 GO terms from each group. In expanded clones, HIV-1 IS were more frequently near genes related to lymphocyte activation and chromatin regulation than in

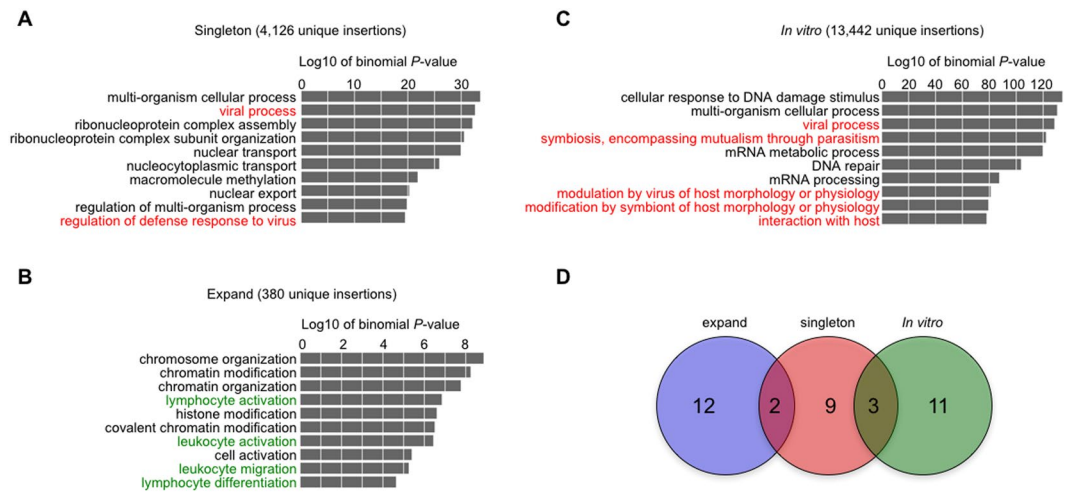


Figure 5. Gene ontology enrichment analysis of HIV-1 integration sites. (A,B) GO enrichment analysis of the HIV-1 integration sites in singleton (A) and expanded clones (B). (C) The GO enrichment analysis for *in vitro* integration sites is shown as a control. The numbers above the graphs represent $-\log_{10}$ of binomial *P*-value. There were 69, 14, and 105 GO terms enriched significantly in singleton clones, expanded ones, and *in vitro* infection, respectively. Term names shown in red and green are associated with “Host-pathogen interaction” and “leukocyte and lymphocyte activation”, respectively. (D) Venn diagrams showing the similarities and differences of enriched GO terms between each data set. For comparison, top 14 term names were selected from each data set.

GO_Id	GO_name	Odds ratio	<i>P</i> -value	Q-value
GO:0050900	leukocyte migration	0.522	0.008	0.059
GO:0016570	histone modification	0.568	0.009	0.059
GO:0051276	chromosome organization	0.664	0.009	0.059
GO:0016568	chromatin modification	0.635	0.009	0.059
GO:0016569	covalent chromatin modification	0.571	0.009	0.059
GO:0006325	chromatin organization	0.651	0.012	0.065
GO:0045321	leukocyte activation	0.643	0.019	0.079
GO:0046649	lymphocyte activation	0.627	0.018	0.079
GO:0001775	cell activation	0.702	0.040	0.148

Table 1. List of GO terms enriched in expanded clones. GO terms included in the Venn diagrams (Fig. 5D) were analyzed to determine statistical significance. *P*-value and Q-value was determined by using Fisher’s exact test or Storey Tibshirani method⁵², respectively. Odds ratio was calculated as $S_p E_A / S_A E_p$ (S_p : number of singleton IS present nearby genes related to a specific GO term, S_A : number of singleton IS absent nearby those genes, E_p : number of expanded IS present nearby those genes, and E_A : number of expanded IS absent nearby those genes).

singleton clones, which reached statistical significance (Table 1). These results indicated potential drivers of clonal expansion of HIV-1-infected cells *in vivo*.

Some expanded clones were detectable in multiple organs. Almost all previous studies on the clonality of HIV-1 infected cells focused on CD4⁺ T cells or mononuclear cells in peripheral blood, except for one case report¹⁸. In order to analyze whether the expansion of HIV-1-infected cells in the humanized mouse is systemic or a local phenomenon, we compared the data on integration sites among different organs in a 15 wpi mice (mouse no. 5, Fig. 6A). Because of the high sensitivity of the high-throughput sequencing and analysis pipeline, there is a risk of technical false positives, due to an overflow of highly abundant sequence reads into other samples (Supplementary Fig. S1), which could lead to an overestimation of the extent of systemic expansion of HIV-1-infected cells. To minimize the risk of such false positive results, we used strict criteria to evaluate clonal expansion. We analyzed multiple samples in the same run and then separated the data by the sequence of index reads (Supplementary Fig. S1). The multiplexing step can also generate contamination of sequence reads across samples. To reduce the possibility of errors caused by the multiplexing, we filtered the data based on the quality of the index reads. We utilized 8-bp index reads and selected high-quality reads (Phred quality score > 30) only. In theory, the risk of misidentification of an index was less than 1×10^{-15} . In addition, we discarded sequence reads that contained more than one low quality residue (Phred quality score < 20) in the Read 1 and Read 2 data, to reduce the possibility of false positives due to sequencing or mapping errors. These strict criteria enabled

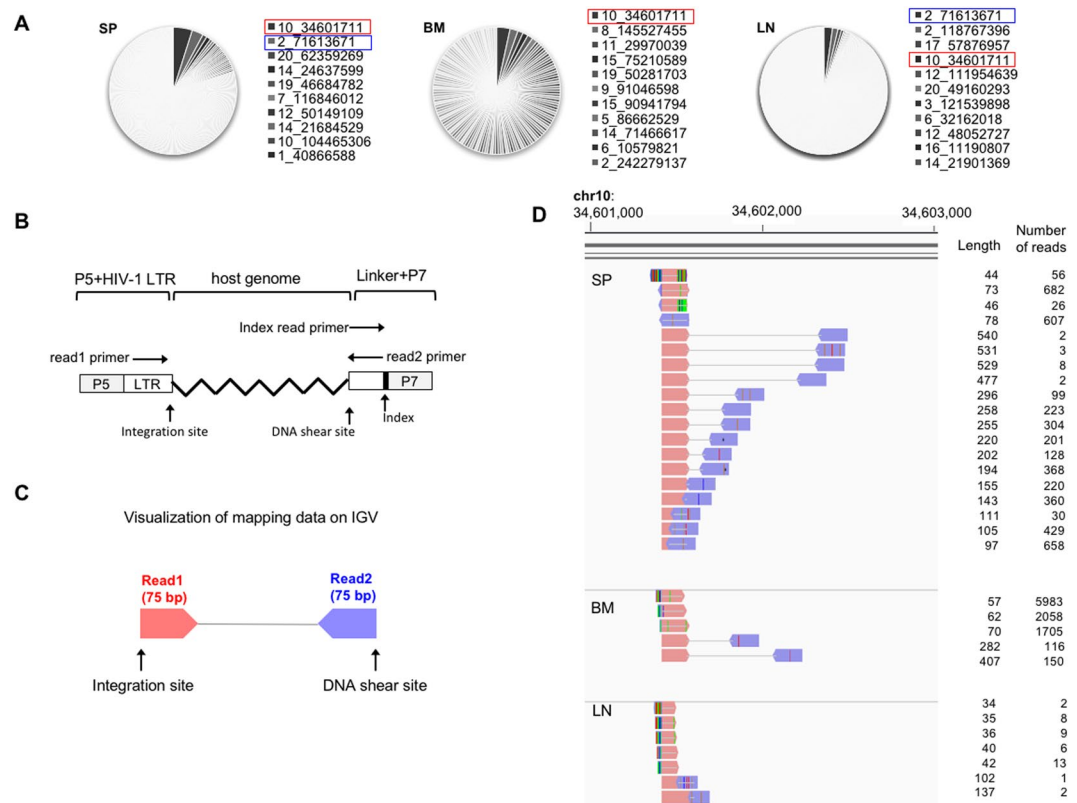


Figure 6. Evidence of systemic expansion of HIV-1-infected clones. **(A)** Pie charts of the abundance of HIV-1-infected clones in the mouse no. 5 shown in Fig. 3B. Red rectangles highlight a highly expanded clone detected in SP, BM, and LN. Blue rectangles highlight a highly expanded clone in SP and LN (Supplementary Fig. S7). **(B)** Schematic diagram to show the positions of each primer used for Illumina sequencing. **(C)** Schematic diagram for visualizing mapped data on IGV software. **(D)** Visualization of sequencing data of an expanded clone detected in all tissues analyzed. The length between the integration site and the DNA shear site is shown on the right side of each amplicon. The number of reads is also shown on the right side of each amplicon.

us to minimize the chance of overestimation of the frequency of abundant clones. The results from a 15-wpi mouse showed that several expanded clones were detectable in multiple tissues (Fig. 6A). For example, an HIV-1-infected expanded clone with the integration site in chr10: 34,601,711 was detected in SP, BM and LN. In each such clone, the integration site was identical, but the DNA shear site was different among those tissues, suggesting that the detection of the same clone in different tissue was not the result of cross-contamination but rather represented true systemic expansion of the HIV-1 infected clone (Fig. 6B and D). Using the same strict criteria, other expanded clones were detected in both SP and LN (Supplementary Fig. S7) and also in other 15 wpi mice (Supplementary Fig. S8).

Discussion

The high-throughput sequencing of randomly-sheared DNA has made it possible to comprehensively quantify the clonality of retrovirus-infected cells^{16,17}, and the approach has led to recent reports of clonal expansion of HIV-1-infected cells in the peripheral blood of infected individuals^{12,14,15}. These important findings raise the question whether clonally expanded cells contribute to the latent viral reservoir. For instance, a previous report suggested that clonally expanded HIV-1-infected cells might not contribute to maintain the viral reservoir because all expanded clones appeared to contain defective HIV-1 proviruses¹². In contrast, another study showed that expanded clones that accumulate near cancer tissue in HIV-1-infected patients have the capacity to produce infectious viral particles¹⁸. A recent comprehensive study by using PBMCs and LNs in individuals with natural virologic control demonstrated that some expanded clones could induce virion production after T-cell stimulation¹³. These data suggest that at least some expanded HIV-1 infected clones can function as a viral reservoir.

Several critical issues remain to be addressed regarding clonally expanded HIV-1-infected cells, such as the kinetics of their emergence and their tissue distribution. The findings reported in this study suggest that clonal expansion of HIV-1 infected cells is rapid and happens within two weeks of infection. Furthermore, such expanded clones are distributed throughout the lymphoid organs in our humanized mouse model. Their survival and expansion is remarkable in the face of a steep drop in the total number of CD4⁺ T cells in 15 wpi mice (Fig. 1C), suggesting a strong selective advantage. Since the population size and type of CD4⁺ T-cell subsets at 15 wpi can be changed compared to those at 2 wpi, it is quite plausible that the estimates of infected cells types (clonal and singleton) can be impacted by these effects. This possibility will be addressed in future investigations.

Also, it will be interesting to investigate if some of these expanded clones bear a functional, yet latent provirus and if they belong to specific classes of memory CD4⁺ T cells^{27,28}.

Another critical question that remains to be addressed is the mechanism that causes clonal expansion of HIV-1 infected cells. Previous studies demonstrated frequent integration into cell growth and cancer-related genes in expanded clones^{14,15,29}, thus HIV-1 integration might induce aberrant expression of cancer-related genes, resulting in enhanced cell proliferation. Alternatively, some HIV-1 infected cells might have a higher proliferative capacity, independently of any effect of the integrated provirus. In this study, we observed frequent HIV-1 IS into genes associated with lymphocyte activation, proliferation, and chromatin regulation in expanded clones, but less so in singletons. Activated, proliferating CD4⁺ T cells are characterized by a high expression of the genes related with cell activation. Because HIV-1 preferentially integrates into highly transcribed genes^{9,12}, a scenario is plausible whereby a significant number of HIV-1 infections occurs in actively proliferating CD4⁺ T cells, resulting in frequent proviral integration near or within genes associated with cell activation. These rapidly proliferating CD4⁺ T cells likely give rise to expanded clones, some of which may be long-lived if the provirus is defective or later becomes latent, for example by transcriptional interference^{30,31}. This picture is consistent with the observation that HIV-1 persists in memory T-cells, which undergo mitosis more frequently than naïve T cells^{13,32}.

As we have shown in this study, the humanized mouse model should be useful to address key questions on clonal expansion of HIV-1 infected cells for the following reasons:

First, all tissues are accessible in the mice, in addition to PBMCs. Most of previous studies were limited to PBMCs yet HIV-1 infection involves tissues such as the spleen, lymph nodes, the gastrointestinal mucosa and the central nervous system³³. We show here that the clonal expansion of HIV-1-infected cells is not a local phenomenon but systemic because we observed the same expanded infected clone in multiple organs (Fig. 6). We found that clonal expansion was more evident in BM than in SP or LN (Fig. 2C and D). The frequency of integration into host genes was significantly higher in LN than other tissues (Fig. 4C) and it was close to that in cells infected *in vitro* (Fig. 4A), suggesting that *de novo* infection of HIV-1 in LN is more evident than other tissues. This idea is in line with findings in rhesus macaque infected with simian immunodeficiency virus and HIV-1 infected individuals with natural virologic control, where there is persistent viral replication in both follicular and non-follicular helper T cells in lymph nodes^{13,34}. Further studies are needed to understand the organ-specific dynamics of clonal expansion of HIV-1 infected cells.

Second, there is less heterogeneity in the humanized mice than in infected individuals. Previous studies identified thousands of integration sites in human subjects, but the number of patients analyzed was still small. Differences in terms of founder virus sequence, route of initial infection, duration of infection, treatment regimen and duration are likely to affect the clonal expansion of HIV-1 infected cells and their tissue distribution. To identify the fundamental underlying mechanisms of clonal expansion of infected cells, *in vivo* models with less heterogeneity should be advantageous. In this study, we demonstrated that clonal expansion occurred within 2 weeks of infection (Fig. 2A). In line with a previous report in human subjects¹², the degree of clonal expansion of infected cells in the humanized mice increased with the duration of infection from 2 wpi (16.7%) to 15 wpi (25.8%) (Fig. 2A). In the mouse model, we will be able to analyze the kinetics of clonal expansion with greater confidence, because we will know the precise time of infection.

Third, the spread of HIV-1 in the humanized mouse occurs in the situation that the host immune response against virus is incompetent: the results may therefore mimic the spread of HIV-1 in early infection in the human, in whom an effective immune response to the virus is typically mounted after 1 to 4 months³⁵. Because of the relatively short period of infection and lower selection pressure exerted by the host immune system, sequence variation of HIV-1 in the humanized mice should be much lower than that in human subjects³⁶, reducing the frequency of undetected clones. Due to the design of the oligonucleotide primers used to analyze the integration sites, sequence variation in the primer-binding region of the provirus may preclude detection of certain clones. The fact that we may miss some clones in the integration site analysis must be taken into consideration when interpreting the clonality data. Several alternative methods for IS analysis to overcome this point are being developed^{37,38}.

In summary, we have shown rapid generation of a large number of clones of HIV-1-infected cells and expansion of individual clones in the humanized mouse. Individual expanded clones have disseminated into multiple organs at 15 weeks post infection. The HIV-1 provirus in expanded clones was integrated significantly more often near host genes with specific gene ontology such as lymphocyte activation and chromatin regulation, compared with the singleton clones. Although it still remains unknown whether HIV-1 integration directly causes clonal expansion of the infected host cells, our analyses of 18,104 integration sites from different data sets both *in vitro* and *in vivo* provide concrete evidence for the association between HIV-1 IS and the clonal expansion of HIV-1-infected cells and strengthened the application of the humanized mouse system to elucidate further mechanistic insights into persistent HIV-1 infection *in vivo*.

Methods

Ethics statement. All procedures including animal studies were conducted following the guidelines for the Care and Use of Laboratory Animals of the Ministry of Education, Culture, Sports, Science and Technology, Japan. The authors received approval from the Institutional Animal Care and Use Committees (IACUC)/ethics committee of Kyoto University institutional review board (protocol number D13–25). All protocols involving human subjects were reviewed and approved by the Kyoto University institutional review board. The study was carried out in accordance with the guidelines proposed in the Declaration of Helsinki. Informed written consent from human subjects was obtained in this study.

HIV-1 infection to humanized mice. NOD/SCID/*Il2rg*^{null} mice (NOG mice) were obtained from the Central Institute for Experimental Animals (Kawasaki, Japan)³⁹. Human CD34⁺ hematopoietic stem cells were isolated from human fetal liver kindly provided by Dr. Dong Sung An (University of Los Angeles, USA). We

generated humanized mice (NOG-hCD34 mice) as described previously^{40,41}. HIV-1_{JR-CSF} viral solution was prepared as previously described⁴². Briefly, pJR-CSF, an infectious molecular clone of HIV-1_{JR-CSF}^{43,44}, was transfected into 293T cells. At 48 h post-transfection, the culture supernatant was harvested, centrifuged, and then filtered through a 0.45- μ m-pore-size filter to produce virus solutions. Viral solution (1,500 infectious units [IU]) was intraperitoneally injected into the NOG-hCD34 mice. All mice were maintained in accordance with institutional guidelines for animal welfare at Kyoto University. Peripheral blood and plasma samples were collected at 0, 1, 2, 3, 5, 7, 9, 12, and 15 wpi as previously described^{23,36,41,45,46}. The amount of HIV-1 RNA in 50 μ L plasma was quantified by Bio Medical Laboratories, Inc. (Tokyo, Japan). CD4⁺ T-cell numbers in peripheral blood were calculated by flowcytometry using anti-CD4 antibody (RPA-T4; Biolegend). The mice were sacrificed at 2 or 15 wpi under anesthesia, and then human mononuclear cells (hMNCs) of the LN, SP, and BM of humanized mice were isolated and purified as previously described^{23,36,41,45,46}. LNs were gently homogenized using a homogenizer pestle. SPs were crushed and rubbed on a steel mesh with 1-mm grids to generate single cell suspensions in RPMI1640 supplemented with 4% FCS. To collect BM, femurs were dissected at both ends and the interior was flushed with RPMI 640 supplemented with 4% FCS. hMNCs in the spleen and bone marrow of humanized mice were further purified from the MNCs of these organs using Ficoll-Paque (Pharmacia). The human MNCs were stored in Cell Banker (Juji Field Inc., Tokyo, Japan) at -80°C until use.

Preparation of HIV-1 or HTLV-1-infected cells *in vitro*. To prepare HIV-1-infected cells *in vitro* Jurkat E6.1 (ECACC) cells were infected cells with recombinant HIV-1 *in vitro* as described previously⁴⁷. Briefly, Jurkat cells (30 mL at $0.8 \times 10^6/\text{mL}$) were infected with VSV-G pseudotyped LAI Δ env at an MOI of 0.2. An aliquot of cells was analysed by FACS 36 hours post-infection and DNA was extracted from the remaining cells using Qiagen DNeasy Blood & Tissue kit following the manufacturer's instructions. We also prepared HTLV-1 infected cells *in vitro* as describe before¹⁶. Briefly, Jurkat E6.1 (ECACC) cells were infected with HTLV-1 by co-cultivation with MT-2, a virus-producing cell line. To reduce contamination with MT-2 cells, we treated MT-2 with γ -irradiation before co-cultivation. DNAs were extracted from Jurkat cells infected with HTLV-1 using DNeasy Blood and Tissue kit (Qiagen).

Linker-mediated (LM)-PCR. Integration site analysis of HIV-1 was performed using linker-mediated PCR and high-throughput sequencing as described previously with minor modifications¹⁶, which is the same strategy as a previous report analyzing PBMCs of HIV-1-infected individuals¹⁴. Briefly, 1 microgram genomic DNA was fragmented by sonication with a Covaris S220 (Covaris, Inc., MA) or Picoruptor (Diagenode, S.A., Belgium) device to produce fragments in the range of 300–500bp. The DNA ends were repaired and a DNA linker was added. The junction between the 3'/LTR of HIV-1 and the host genomic DNA was amplified with a primer targeting the 3'/LTR and another targeting the linker. After a nested PCR, PCR amplicons were quantified using Illumina P5 and P7 primers. A list of oligonucleotides used in the LM-PCR is given in Supplementary Table S4. DNA libraries were sequenced by Illumina MiSeq as paired-end reads, and analysis performed on the resulting fastq files.

Proviral load (PVL) measurement. We estimated the number of infected cells by quantifying the copy number of the *gag* gene, normalized to the copy number of the *ALB* gene, according to a previous report, with minor modification²⁵. The presence of unintegrated 2-LTR circles⁴⁸ in infected cells can contribute to the copy number of *gag* in the assay. Therefore we subtracted the copy number of the 2-LTR DNA from the total copy number of *gag*. The proviral load was calculated as follows: PVL (%) = ((copy number of *gag*) – (copy number of 2-LTR))/((copy number of albumin)/2) \times 100. Primer sequences are listed in Supplementary Table S5. In this PVL measurement, proviruses defective of *gag* region and the presence of unintegrated 1-LTR circles could affect the PVL value. Thus, we used “estimated PVL” as the y-axis label of proviral load measurement (Fig. 1C and D).

RNA-seq. To obtain the human CD4⁺ T cell fraction (CD45⁺CD3⁺CD8[−]) from the humanized mice, human splenic MNCs were stained with PE-conjugated anti-CD45-PE (HI30; Biolegend), APC-Cy7-conjugated anti-CD3 (HIT3a; Biolegend), and PE-Cy7-conjugated anti-CD8 antibody (HIT8a; Biolegend), and CD45⁺CD3⁺CD8[−] cells were sorted with FACSJazz (BD Biosciences). The purity was >99%. RNA was extracted from CD4⁺ splenocytes using RNeasy (Qiagen). cDNA libraries were generated using TrueSeq RNA Access Library Prep Kit (Illumina). Sequencing was performed using TruSeq SR Cluster Kit v3 and TruSeq SBS Kit v3 – HS (50 cycle) (Illumina) by Medical Biological Laboratories Co., Ltd (Nagoya, Japan). HiSeq. 2000 (Illumina) generated fastq files with around 40–80 million reads per sample. RNA-seq data analysis was performed using CLC Genomics workbench 7.5 (CLC bio). We generated an average value of RPKM (Reads per kilobase of exon model per million mapped reads) for each human gene, and used these values to classify genes into 4 groups (no, low, medium, or high expression).

High-throughput sequencing data analysis. Each cluster on the Illumina flow cell generated 3 fastq files, including Read1, Read2, and Index Read. Read1 corresponded to sequencing data generated by a primer within HIV-1 LTR (Fig. 6B) and Read2 to data generated by a primer within the linker. The Index Read corresponded to the 8-bp index sequence in the linker. First, we identified clusters on the flow cell with high sequencing quality of the Index Read (Phred quality score >30 at each position of 8-bp index read) using an R program kindly provided by Michi Miura (Imperial College London, UK). We next removed the linker sequence from Read1 and Read2, and then we selected reads with LTR sequence in Read1 (TGACA for HIV-1). After trimming the 5-bp viral sequence from Read1, we performed a further sequencing quality check using a bespoke Perl script (Amelieff, Tokyo, Japan). This software processes sequence data as follows: i) removes reads containing more than 80% bases with Q score less than 20, ii) removes reads less than 20 bp in length, and iii) removes unpaired reads. The cleaned sequencing reads were mapped to the HIV-1 genomes, (Genbank, K03455) with or without

the human reference genome (hg19) using the BWA-MEM algorithm⁴⁹. We used Samtools for further data processing, such as removal of multiple mapping to the reference genome⁵⁰. Since the LTR sequence is identical, the mapping data contained information of both 5'- and 3'-LTR. To identify integration sites, we removed data from the 5'-LTR. For the visualization of the mapped results as shown in Fig. 6C, we used Integrative Genomics Viewer (IGV) (<http://software.broadinstitute.org/software/igv/>). For clonality analysis of HIV-1-infected cells, we exported the files containing the information on integration sites, DNA shear sites, and number of reads (Fig. 6B). We then calculated the number of copies of each individual clone by counting the number of different shear sites present for each integration site¹⁷. There are two possibilities when we find two cells with the same IS. One possibility is that we are detecting the clonal expansion of one original clone. The other is HIV-1 integrating into the same integration site of two different host cells. The former possibility is far more likely than the latter. Because the human genome contains about 3 billion base pairs, the possibility of finding the same integration site by chance in two different clones is 0.00000003%. In the current study, we identified 4,662 unique integration sites in humanized mice. Therefore, the frequency that two different infected clones happen to have the same integration site is far less than 0.001%. Furthermore, the result shown in Fig. 2A, in which there were far less expanded clones *in vitro* infection than *in vivo*, strongly supports the former possibility. Thus, when we observed two cells with same integration site, we defined them as expanded clones in this study.

Bioinformatic analysis. We generated a bed file from the exported file containing the information on integration sites. RefSeq gene data was obtained from UCSC tables (<https://genome.ucsc.edu/>). Positions of RefSeq genes were compared to the IS using the R package hiAnnotator (<http://github.com/malnirav/hiAnnotator>). We also performed Gene ontology (GO) analysis using GREAT, an online software application for gene annotations⁵¹ (<http://bejerano.stanford.edu/great/public/html/index.php>).

Statistical analysis. Data were analyzed using a chi-squared test with Prism 7 software (GraphPad Software, Inc., CA) unless otherwise described. Statistical significance was defined as $P < 0.05$.

Data availability. Fastq files obtained in this study have been deposited in the DNA Data Bank of Japan (DDBJ). (accession no. DRA005133). Data on the findings reported here are available from the corresponding author upon request.

References

1. Palella, F. J. Jr. *et al.* Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *N Engl J Med* **338**, 853–860, doi:10.1056/NEJM199803263381301 (1998).
2. Deeks, S. G. *et al.* Towards an HIV cure: a global scientific strategy. *Nat Rev Immunol* **12**, 607–614, doi:10.1038/nri3262 (2012).
3. Palmer, S., Josefsson, L. & Coffin, J. M. HIV reservoirs and the possibility of a cure for HIV infection. *J Intern Med* **270**, 550–560, doi:10.1111/j.1365-2796.2011.02457.x (2011).
4. Antiretroviral Therapy Cohort, C. Causes of death in HIV-1-infected patients treated with antiretroviral therapy, 1996–2006: collaborative analysis of 13 HIV cohort studies. *Clin Infect Dis* **50**, 1387–1396, doi:10.1086/652283 (2010).
5. Hasse, B. *et al.* Morbidity and aging in HIV-infected persons: the Swiss HIV cohort study. *Clin Infect Dis* **53**, 1130–1139, doi:10.1093/cid/cir626 (2011).
6. Ho, Y. C. *et al.* Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540–551, doi:10.1016/j.cell.2013.09.020 (2013).
7. Archin, N. M., Sung, J. M., Garrido, C., Soriano-Sarabia, N. & Margolis, D. M. Eradicating HIV-1 infection: seeking to clear a persistent pathogen. *Nat Rev Microbiol* **12**, 750–764, doi:10.1038/nrmicro3352 (2014).
8. Delagrèverie, H. M., Delaugerre, C., Lewin, S. R., Deeks, S. G. & Li, J. Z. Ongoing Clinical Trials of Human Immunodeficiency Virus Latency-Reversing and Immunomodulatory Agents. *Open Forum Infect Dis* **3**, ofw189, doi:10.1093/ofid/ofw189 (2016).
9. Schroder, A. R. *et al.* HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**, 521–529 (2002).
10. Singh, P. K. *et al.* LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. *Genes Dev* **29**, 2287–2297, doi:10.1101/gad.267609.115 (2015).
11. Bruner, K. M. *et al.* Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med* **22**, 1043–1049, doi:10.1038/nm.4156 (2016).
12. Cohn, L. B. *et al.* HIV-1 integration landscape during latent and active infection. *Cell* **160**, 420–432, doi:10.1016/j.cell.2015.01.020 (2015).
13. Boritz, E. A. *et al.* Multiple Origins of Virus Persistence during Natural Control of HIV Infection. *Cell* **166**, 1004–1015, doi:10.1016/j.cell.2016.06.039 (2016).
14. Maldarelli, F. *et al.* HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179–183, doi:10.1126/science.1254194 (2014).
15. Wagner, T. A. *et al.* HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* **345**, 570–573, doi:10.1126/science.1256304 (2014).
16. Gillet, N. A. *et al.* The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones. *Blood* **117**, 3113–3122, doi:10.1182/blood-2010-10-312926 (2011).
17. Berry, C. C. *et al.* Estimating abundances of retroviral insertion sites from DNA fragment length data. *Bioinformatics* **28**, 755–762, doi:10.1093/bioinformatics/bts004 (2012).
18. Simonetti, F. R. *et al.* Clonally expanded CD4+ T cells can produce infectious HIV-1 *in vivo*. *Proc Natl Acad Sci U S A* **113**, 1883–1888, doi:10.1073/pnas.1522675113 (2016).
19. Margolis, D. M., Garcia, J. V., Hazuda, D. J. & Haynes, B. F. Latency reversal and viral clearance to cure HIV-1. *Science* **353**, aaf6517, doi:10.1126/science.aaf6517 (2016).
20. Koyanagi, Y., Tanaka, Y., Ito, M. & Yamamoto, N. Humanized mice for human retrovirus infection. *Curr Top Microbiol Immunol* **324**, 133–148 (2008).
21. Legrand, N. *et al.* Humanized mice for modeling human infectious disease: challenges, progress, and outlook. *Cell Host Microbe* **6**, 5–9, doi:10.1016/j.chom.2009.06.006 (2009).
22. Kaneshima, H. *et al.* Human immunodeficiency virus infection of human lymph nodes in the SCID-hu mouse. *Proc Natl Acad Sci USA* **88**, 4523–4527 (1991).
23. Nie, C. *et al.* Selective infection of CD4+ effector memory T lymphocytes leads to preferential depletion of memory T lymphocytes in R5 HIV-1-infected humanized NOD/SCID/IL-2R γ manull mice. *Virology* **394**, 64–72, doi:10.1016/j.virol.2009.08.011 (2009).

24. Sato, K. *et al.* Comparative study on the effect of human BST-2/Tetherin on HIV-1 release in cells of various species. *Retrovirology* **6**, 53, doi:10.1186/1742-4690-6-53 (2009).
25. Douek, D. C. *et al.* HIV preferentially infects HIV-specific CD4+ T cells. *Nature* **417**, 95–98, doi:10.1038/417095a (2002).
26. Brady, T. *et al.* HIV integration site distributions in resting and activated CD4+ T cells infected in culture. *AIDS* **23**, 1461–1471, doi:10.1097/QAD.0b013e32832caf28 (2009).
27. Fromentin, R. *et al.* CD4+ T Cells Expressing PD-1, TIGIT and LAG-3 Contribute to HIV Persistence during ART. *PLoS Pathog* **12**, e1005761, doi:10.1371/journal.ppat.1005761 (2016).
28. Chomont, N. *et al.* HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat Med* **15**, 893–900, doi:10.1038/nm.1972 (2009).
29. Ikeda, T., Shibata, J., Yoshimura, K., Koito, A. & Matsushita, S. Recurrent HIV-1 integration at the BACH2 locus in resting CD4+ T cell populations during effective highly active antiretroviral therapy. *J Infect Dis* **195**, 716–725, doi:10.1086/510915 (2007).
30. Han, Y. *et al.* Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. *Cell Host Microbe* **4**, 134–146, doi:10.1016/j.chom.2008.06.008 (2008).
31. Lenasi, T., Contreras, X. & Peterlin, B. M. Transcriptional interference antagonizes proviral gene expression to promote HIV latency. *Cell Host Microbe* **4**, 123–133, doi:10.1016/j.chom.2008.05.016 (2008).
32. Murray, A. J., Kwon, K. J., Farber, D. L. & Siliciano, R. F. The Latent Reservoir for HIV-1: How Immunologic Memory and Clonal Expansion Contribute to HIV-1 Persistence. *J Immunol* **197**, 407–417, doi:10.4049/jimmunol.1600343 (2016).
33. Ruelas, D. S. & Greene, W. C. An integrated overview of HIV-1 latency. *Cell* **155**, 519–529, doi:10.1016/j.cell.2013.09.044 (2013).
34. Fukazawa, Y. *et al.* B cell follicle sanctuary permits persistent productive simian immunodeficiency virus infection in elite controllers. *Nat Med* **21**, 132–139, doi:10.1038/nm.3781 (2015).
35. Koup, R. A. *et al.* Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *J Virol* **68**, 4650–4655 (1994).
36. Sato, K. *et al.* Remarkable lethal G-to-A mutations in vif-proficient HIV-1 provirus by individual APOBEC3 proteins in humanized mice. *J Virol* **84**, 9546–9556, doi:10.1128/JVI.00823-10 (2010).
37. Sunshine, S. *et al.* HIV Integration Site Analysis of Cellular Models of HIV Latency with a Probe-Enriched Next-Generation Sequencing Assay. *J Virol* **90**, 4511–4519, doi:10.1128/JVI.01617-15 (2016).
38. Miyazato, P. *et al.* Application of targeted enrichment to next-generation sequencing of retroviruses integrated into the host human genome. *Sci Rep* **6**, 28324, doi:10.1038/srep28324 (2016).
39. Ito, M. *et al.* NOD/SCID/gamma(c)(null) mouse: an excellent recipient mouse model for engraftment of human cells. *Blood* **100**, 3175–3182, doi:10.1182/blood-2001-12-0207 (2002).
40. Sato, K. *et al.* HIV-1 Vpr accelerates viral replication during acute infection by exploitation of proliferating CD4+ T cells *in vivo*. *PLoS Pathog* **9**, e1003812, doi:10.1371/journal.ppat.1003812 (2013).
41. Sato, K. *et al.* APOBEC3D and APOBEC3F potentially promote HIV-1 diversification and evolution in humanized mouse model. *PLoS Pathog* **10**, e1004453, doi:10.1371/journal.ppat.1004453 (2014).
42. Sato, K. *et al.* Modulation of human immunodeficiency virus type 1 infectivity through incorporation of tetraspanin proteins. *J Virol* **82**, 1021–1033, doi:10.1128/JVI.01044-07 (2008).
43. Kawano, Y. *et al.* Mutational analysis of human immunodeficiency virus type 1 (HIV-1) accessory genes: requirement of a site in the nef gene for HIV-1 replication in activated CD4+ T cells *in vitro* and *in vivo*. *J Virol* **71**, 8456–8466 (1997).
44. Koyanagi, Y. *et al.* Dual infection of the central nervous system by AIDS viruses with distinct cellular tropisms. *Science* **236**, 819–822 (1987).
45. Sato, K. *et al.* A novel animal model of Epstein-Barr virus-associated hemophagocytic lymphohistiocytosis in humanized mice. *Blood* **117**, 5663–5673, doi:10.1182/blood-2010-09-305979 (2011).
46. Sato, K. *et al.* Vpu augments the initial burst phase of HIV-1 propagation and downregulates BST2 and CD4 in humanized mice. *J Virol* **86**, 5000–5013, doi:10.1128/JVI.07062-11 (2012).
47. Zhou, L. *et al.* Transportin 3 promotes a nuclear maturation step required for efficient HIV-1 integration. *PLoS Pathog* **7**, e1002194, doi:10.1371/journal.ppat.1002194 (2011).
48. Pauza, C. D., Trivedi, P., McKechnie, T. S., Richman, D. D. & Graziano, F. M. 2-LTR circular viral DNA as a marker for human immunodeficiency virus type 1 infection *in vivo*. *Virology* **205**, 470–478, doi:10.1006/viro.1994.1667 (1994).
49. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, doi:10.1093/bioinformatics/btp324 (2009).
50. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079, doi:10.1093/bioinformatics/btp352 (2009).
51. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495–501, doi:10.1038/nbt.1630 (2010).
52. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* **100**, 9440–9445, doi:10.1073/pnas.1530509100 (2003).

Acknowledgements

We thank Michiyo Tokunaga for technical and secretary support, Alex Zhyvoloup for technical support, and Michi Miura for R program to perform quality check of Index reads. We also thank Heather Niederer for providing materials and Laurence Game and Marian Dore (Genomics Laboratory, Medical Research Council Clinical Sciences Centre, Hammersmith Hospital, London, UK) for sequencing of IS obtained from cells infected with HIV-1 or HTLV-1 *in vitro*. This study was supported by grants from the JSPS KAKENHI (26461428, 17K09016 and 15H04870 to Y.S., 16KK0206 to H.K., and 15K07166 to K.S.), AIDS International Collaborative Research Grant from the Ministry of Education, Science, Sports, and Culture, and Kumamoto University to Y.S., AMED (15Afk0410013h0001) to Y.K., CREST, JST to K.S., EU HIVINNOV (305137) to A. Fassati, and the Wellcome Trust (WT100291MA) to C.R.M.B. The funder has no role in the study design of the study, collection of data, its interpretation, or the discussion to submit the work for publication.

Author Contributions

Y.S. conceived the study, designed experiments, analysed and interpreted the data, and wrote the paper; H.K. performed experiment, analysed and interpreted the data, and wrote the paper; A. Fukuda and N.M., performed experiment; J.I. performed bioinformatic analysis; Y.U. performed sequencing data analysis; P.M. and A. Fassati assisted with the experiments and wrote the paper; M.S.I. assisted with the experiments; A.M. and C.R.M.B., assisted with the experiments and analysed and interpreted the data, and wrote the paper; Y.K. provided humanized mice, analysed and interpreted the data, and wrote the paper; and K.S., planned conception, analysed and interpreted the data, and wrote the paper.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-07307-4](https://doi.org/10.1038/s41598-017-07307-4)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017